FYS-STK4155 - Project 2

Geir Tore Ulvik, Idun Klvstad

November 12, 2018

Abstract

1 Introduction

Code, data and figures are available at the following GitHub address: GitHub repository
This project will follow closely the article of Metha et al, arXiv 1803.08823. The article is listed in references: [1].

The main aim of the project is to study both classification and regression problems. The project will include three regression algorithms, logistic regression for classification problems and a multilayer perceptron code. The multilayer perceptron code is for studying both regression and classification problems.

We will use the so-called Ising model for the training data, and also focus on supervised learning.

As mentioned above we will look at both regression and classification problems. Regression will be used to determine the value of the coupling constant of the energy of the one-dimensional Ising model. The classification case will study the one dimensional Ising model in order to classify the phase of the model.

The various algorithms will be seen in comparison. In the end we will summarize and give a critical evaluation of pros and cons. We will look at which algorithm works best for the regression case, and which works best for the classification case. [2]

2 Theory

As the theory behond the three regression methods used as well as error estimates and the bootstrap method is covered in the previous project, we will not restate it here.

2.1 The Ising model

The ising model is a simple binary value system where the variables in the model can take only two values. For examle ± 1 or 0 and 1. [2]

We will look at the physicist's approach, and call the variables for spin. [2]

Given an ensamble of random spin configurations we can assign an energy to each state, using the 1D Ising model with nearest-neighbor interactions:

$$E = -J \sum_{j=1}^{N} S_j S_{j+1} \tag{1}$$

J is the nearest-neighbor spin interaction, and $S_j \epsilon \pm 1$ is a spin variable. N is the chain length. [1] [2]

In one dimension, this model has no phase transitions at finite temperature. [2]

To get a spin model with pairwise interactions between every pair of variables, we choose the following model class:

$$E_{model}[S^i] = -\sum_{j=1}^{N} \sum_{k=1}^{N} J_{j,k} S_j^i S_k^i \tag{2}$$

[1]

In this equation i represents a particular spin configuration. [2]

The goal with this model is to determine the interaction matrix $J_{j,k}$. As the model is linear in $\mathbf{J}$, it is possible to use linear regression.

The problem can be recast on the form

$$E^i_{model} = \mathbf{x}^i \cdot \mathbf{J} \tag{3}$$

2.2 Logistic regression and classification problems

Differently to linear regression, classificaltion problems are concerned with outcomes taking the form of descrete variables. For a specific physical problem, we'd like to identify its state, say whether it is an ordered of disordered system. [3]

Configurations representing states below the critical temperature are calles ordered states, while those above the critical temperature are called disorderes states. [2]

2.3 Accuracy score

2.4 Gradient Decent solver

2.5 Cost functions

3 Method

4 Results

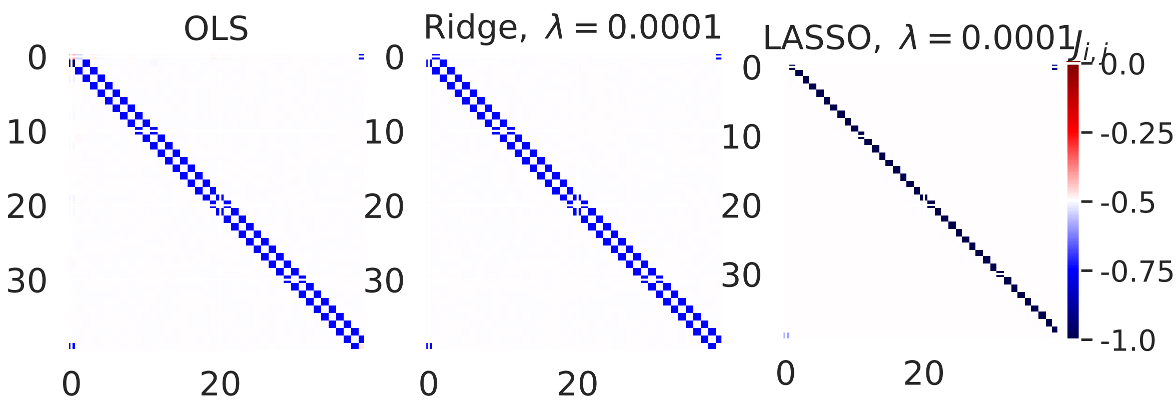

Figure 1: Ising model plots for selected regularization parameters λ. This figure is a work in progress.

Figure 2 show how the R2 score varies between models. It's important to note that the λ for Ridge regression, and α for Lasso regression have the same value, but they affect the models on different orders of magnitude, and must be treated somewhat separately. Even so, we can see that the training and test set R2-scores follow eachother closely. Figure 3 show the results for learning rate η comparison when applying our logistic regression model on the Ising Model data. For some η the model quickly rises to an approximate highest value, but jumps back down to the equivalent of guessing from time to time. When approaching 30 epochs and more, this behaviour seems to diminish somewhat, and overall the etas that produce the best results $(10^{-5} - 10^{-2})$ have most of their values in the higher points. In table 1 the accuracy on data containing critical states is included. Due to the varying nature our logistic model, a solution in which the weights producing the best fit on test data are stored was developed. The accuracies for the critical states are thus not necessarily produced with weights as they are after 30 epochs.

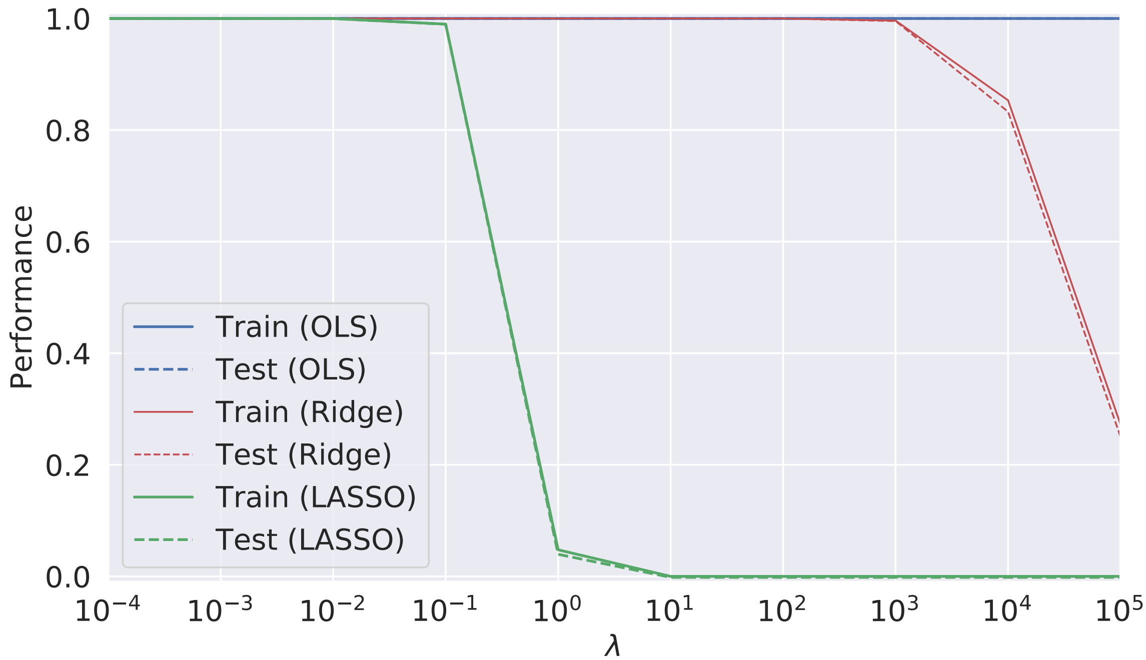

Figure 2: R2 score performance of the linear regression models as a function of regression parameter λ.

η	Training	Test	Critical
10^{-5}	0.718	0.680	0.616
10^{-4}	0.723	0.683	0.624
10^{-3}	0.723	0.685	0.628
10^{-2}	0.712	0.672	0.604
10^{-1}	0.462	0.430	0.460
1	0.465	0.446	0.480

Table 1: Accuracies for a selection of learning rates η after 30 epochs. Batch size = 100, and momentum parameter $\gamma = 0.01$. The ratio of amount of training data to test data is 0.8

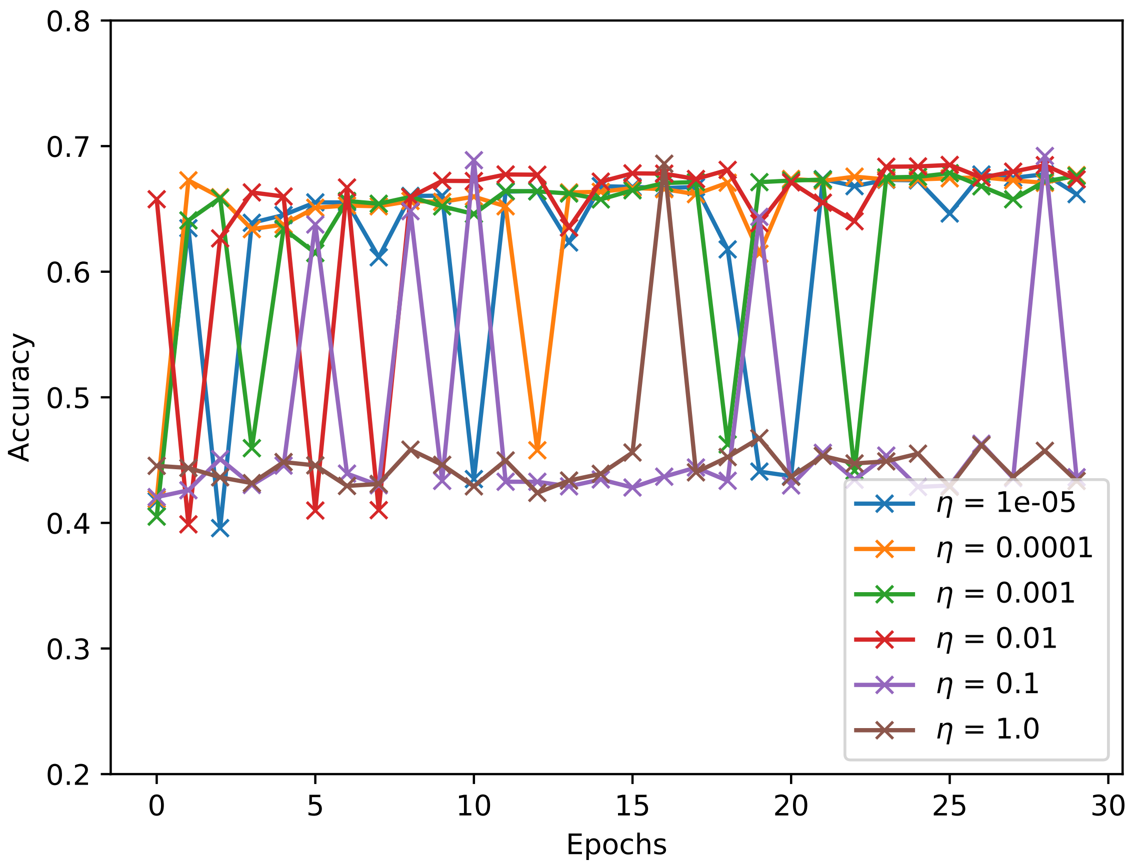

Figure 3: Accuracies for a selection of learning rates η as a function of epochs. 30 epochs, batch size $= 100$, and momentum parameter $\gamma = 0.01$. The accuracy values are measured on the test set. The ratio of amount of training data to test data is 0.8.

5 Discussion

6 Conclusion

References

[1] Pankaj Metha et al. A high-bias, low-variance introduction to machine learning for physicists. 03 2018.

[2] University of Oslo Department of Physics. Project 2 on machine learning. https://compphysics.github.io/MachineLearning/doc/Projects/2018/Project2/pdf/Project2.pdf, 2018.

[3] M Hjorth-Jensen. Data analysis and machine learning: Linear regression and more advanced regression analysis - lecture notes. https://compphysics.github.io/MachineLearning/doc/pub/Regression/html/._Regression-bs000.html, 2018.